

国外著名高等院校
信息科学与技术优秀教材

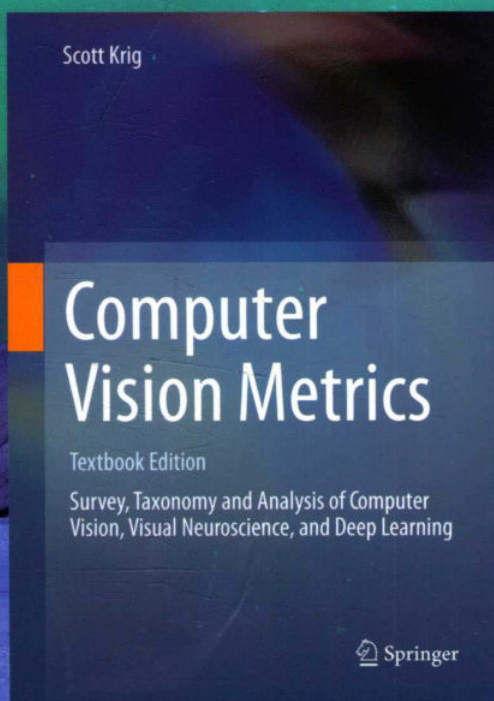
 Springer

 异步图书
www.epubit.com

计算机视觉度量

从特征描述到深度学习

[美] 斯科特·克里格 (Scott Krig) 著 刘波 罗棻 译



 中国工信出版集团

 人民邮电出版社
POSTS & TELECOM PRESS

国外著名高等院校
信息科学与技术优秀教材

 Springer

计算机视觉度量

从特征描述到深度学习

[美] 斯科特·克里格 (Scott Krig) 著 刘波 罗棻 译

人民邮电出版社
北京

图书在版编目 (CIP) 数据

计算机视觉度量：从特征描述到深度学习 / (美)
斯科特·克里格 (Scott Krig) 著；刘波，罗棻译. —
北京：人民邮电出版社，2019.9
国外著名高等院校信息科学与技术优秀教材
ISBN 978-7-115-50588-0

I. ①计… II. ①斯… ②刘… ③罗… III. ①计算机
视觉—高等学校—教材 IV. ①TP302.7

中国版本图书馆CIP数据核字(2019)第016049号

版权声明

Translation from the English language edition
Computer Vision Metrics (Textbook Edition) By Scott Krig
Copyright © Springer International Publishing Switzerland 2016
This work is published by Springer Nature
The registered company is Springer International Publishing AG
All Rights Reserved by the Publisher

本书中文简体字版由施普林格出版社授权人民邮电出版社出版。未经出版者书面许可，不得以任何方式复制或抄袭本书任何部分。

版权所有，侵权必究。

-
- ◆ 著 [美]斯科特·克里格 (Scott Krig)
 - 译 刘波 罗棻
 - 责任编辑 王峰松
 - 责任印制 焦志炜
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号
 - 邮编 100164 电子邮件 315@ptpress.com.cn
 - 网址 <http://www.ptpress.com.cn>
 - 北京鑫正大印刷有限公司印刷
 - ◆ 开本：787×1092 1/16
 - 印张：34.75
 - 字数：827千字 2019年9月第1版
 - 印数：1-2400册 2019年9月北京第1次印刷
 - 著作权合同登记号 图字：01-2017-2289号

定价：148.00元

读者服务热线：(010)81055410 印装质量热线：(010)81055316

反盗版热线：(010)81055315

广告经营许可证：京东工商广登字20170147号

内 容 提 要

本书基于 893 篇参考文献写就，全面介绍了计算机视觉中被广泛使用的各种方法，包括局部特征描述子、区域描述子、全局特征描述子以及评价这些内容的度量方法和分类方法，并用将近一半的篇幅重点介绍了基于深度学习的特征学习方法，以及 FNN、RNN 和 BFN 三类深度学习架构的特点。

本书内容丰富、前沿，强调理论分析，旨在探讨各种计算机视觉研究方法背后的直觉和原理，同时也探讨了深度学习与神经科学之间的关系，指出了未来深度神经网络的发展方向。

本书用专门一章讲解了计算机视觉流程和算法的优化，通过汽车识别、人脸检测、图像分类和增强现实几个实例具体探讨了硬件优化和软件优化的方法。

本书每章末尾都配有相应的思考题，附录给出了许多有效的实践资源和一些有用的分析，同时提供了源代码，既适合高校计算机视觉课程的教学，也适合从事计算机视觉的研究人员和工程技术人员参考使用。

读者在阅读本书时最好具备一定的图像处理和神经网络的知识。

作者简介

Scott Krig 是计算机成像学、计算机视觉和图形可视化方面的先驱。他在 1988 年成立了 Krig Research 公司，该公司提供了世界上第一个基于高性能工程工作站、超级计算机和专有硬件的成像和视觉系统，并为全球 25 个国家的客户提供服务。

近年来，Scott 主要为大型公司和服务于企业市场的初创公司提供服务，帮助它们解决计算机视觉、图形成像、图像学、可视化、机器人、过程控制、工业自动化、计算机安全、密码学，以及成像学和机器视觉在电子消费品（如 PC、笔记本电脑、手机和平板电脑）方面的应用问题。

最近，Scott 将深度感知和计算机视觉方法用于嵌入式系统和移动平台的相关领域，并为 Intel 公司提供技术性指导。Scott 也是全球范围内许多专利应用的发明人，涉及领域包括嵌入式系统、成像学、计算机视觉、DRM 和计算机安全，他也曾在斯坦福大学做过研究工作。

主要译者

刘波，副教授（博士），重庆工商大学计算机科学与信息工程学院教师，主要从事机器学习理论、计算机视觉和最优化技术研究，同时爱好 Hadoop 和 Spark 平台上的大数据分析。

前 言

本书是《计算机视觉度量深入解析》的改进版，增加了一些新的知识点，如深度学习、神经科学在计算机视觉方面的应用，以及神经网络的发展历史和特征学习架构（特别是神经网络方法）等。此外，本书纠正了第1版的拼写错误，增加了3个新章节，同时还介绍了计算机视觉研究中的大量方法，包括局部特征描述子、区域和全局特征、特征学习方法以及组织这些内容的分类法。本书通过详细地分析得出各种研究方法背后的直觉，鼓励读者自己去思考每种方法的动机，创建不同方法的目的，每种方法是如何设计的，以及它们的工作原理是什么。本书还列出了近千篇参考文献，以适合不同层次的读者学习计算机视觉技术。

我对读者的期望是：如果你想学习90%的计算机视觉技术，请阅读本书；还想学习其他10%的技术，请阅读书中列出的参考资料，并至少花上20年的时间创建真实的系统。读完本书需要的时间以小时计算，阅读参考文献并创建真实系统则需花一辈子的时间，本书也只能进行一些肤浅的研究。人们没有时间去了解无关的细节，因此我尽量以一种轻松的方式来介绍计算机视觉知识，使学术界的研究人员和工业界的开发人员都能明白它的基本原理。我很喜欢加勒比海盗 Jack Sparrow 船长的著名格言：“我喜欢它，因为简单易记。”

本书既可作为大学阶段自学参考或课堂教学的书籍，也适合于有经验的工程师和科技人员阅读。根据课程内容的要求，可以采用本书的一些章节来设计出不同的课程。这些课程名可以是：“图像传感器和图像处理”“计算机视觉和图像处理”“计算机视觉的应用和成像优化”“特征学习、深度学习和神经网络架构”“计算机视觉架构”“计算机视觉导论”。每章结尾都提供了相关的练习。建议将本书作为其他优秀计算机视觉书籍、开源代码和实践材料的补充，也可作为高级课程的参考书。

本书不提供大量的编程示例、计算示例、数学证明、实验结果比较，以及详细的性能数据，这些内容在本书参考文献中已经有很好的体现。本书的目标是总结具有代表性的研究方法的成果，而不是重复参考文献中已有的内容。由于本书推荐了很多有效的开源资源给感兴趣的读者，所以本书也不是一本开源代码的工作手册（只提供了少量的源代码）。

相反，本书对计算机视觉研究方法进行了广泛的研究、分类和分析，其目标是找到这些方法背后的原理。读者需要认真阅读本书。它不是一本工作手册，而是为读者提供丰富的技术背景，为各种科学和工程应用寻找进入基础或应用研究的途径。本书完整地介绍了计算机视觉的整个过程，从图像传感器入手，介绍了图像处理、全局特征描述子、区域特征描述子、局部特征描述子、特征学习和深度学习、计算机视觉神经网络、基准数据及训练、跨CPU和GPU的应用工程优化、软件优化方法。目前找不到类似的书籍，这也是我写这本书的动力。在某些方面，这部作品是计算机视觉的一个博物馆，包含概念、观察、新奇的东西和“文物”，让我着迷。

本书旨在从使用特征（特征描述子本身）的角度来对计算机视觉方法进行总结、分类和分析，并介绍这些特征描述子的设计方式和组织方式。学习方法和架构都很重要，本书包含了这两方面的内容。但我对特征描述子很着迷，这些方法以艺术的形式，用数学的方式排列像素图案、形状和光谱，从而揭示了图像的生成方法。我将每个特征描述子都视为艺术作品（如艺术家准备的颜料），因此本书的目的是概述特征描述子和特征学习方法，并欣赏它们。

本书反复强调研究人员需了解特征描述子的有效性，并且要明白起决定性作用的是特征层次结构中使用的绝对特征数量，而不是究竟该如何在 SIFT、像素块和 CNN 特征之间进行选择。本书介绍了很多有用的学习方法、训练方法和架构，并且认为分层特征学习是当前计算机视觉的中流砥柱。计算机视觉从卷积神经网络的开创性工作开始，近年来得到了深度学习方法的快速推动。旧的计算机视觉方法与新方法的结合，使得相关的应用程序开始出现在消费产品中。

特别感谢 Springer 的 Courtney Clarke 委托我完成本书，为了做得更好，他们还提供了相应的支持和指导。

特别感谢大家对第 1 版的精彩反馈，这有助于更好地完成本书。嵌入式视觉联盟 (EVA) 的 Vin Ratford 和 Jeff Bier 为所有 EVA 成员提供第 1 版的副本，包括纸制版和电子版，并保留了一个反馈网页以供大家评论。感谢 Mike Schmidt 和 Vadim Pisarevsky 对本书的出色评论。Juergen Schmidhuber 提供了神经网络和其他有用信息的历史信息链接，Kunihiko Fukushima 给出了他早期神经网络研究论文的副本，Rahul Suthankar 更新了计算机视觉关键趋势，Hugo LaRochelle 和 Patrick Cox 分别提供了有关 CNN 和 HMAX 主题的信息和参考文献。Robert Gens、Andrej Karpathy 也提供了一些有趣的信息。我想再次感谢那些为第 1 版做出贡献的人，他们是 Paul Rosin（关于合成兴趣点），Yann LeCun（为深度学习和卷积神经网络提供了重要参考），Shree Nayar（允许使用多幅图像），Luciano Ovidio（关于蓝天的讨论）。还要感谢影响我想法的人们，他们是 Alexandre Alahi、Steve Seitz、Bryan Russel、Liefeng Bo、Xiaofeng Ren、Gutemberg Guerra-filho、Harsha Viswana、Dale Hitt、Joshua Gleason、Noah Snavely、Daniel Scharstein、Thomas Salmon、Richard Baraniuk、Carl Vondrick、Herve' Jegou、Andrew Richardson、Ofri Weschler、Hong Jiang、Andy Kuzma、Michael Jeronimo、Eli Turiel，以及其他许多没有提到的人。

像往常一样，感谢我的妻子对我的忍耐，感谢她的“限速器开关”调整我工作的节奏，否则我可能会很快耗尽自己的体力。最重要的是，特别感谢那些一直激励着我们的伟大发明家。

Scott Krig

资源与支持

本书由异步社区出品，社区（<https://www.epubit.com/>）为您提供相关资源和后续服务。

配套资源

本书提供如下资源：

- 书中彩图文件。

要获得以上配套资源，请在异步社区本书页面中单击 **配套资源**，跳转到下载界面，按提示进行操作即可。注意：为保证购书读者的权益，该操作会给出相关提示，要求输入提取码进行验证。

提交勘误

作者和编辑尽最大努力来确保书中内容的准确性，但难免会存在疏漏。欢迎您将发现的问题反馈给我们，帮助我们提升图书的质量。

当您发现错误时，请登录异步社区，按书名搜索，进入本书页面，单击“提交勘误”，输入勘误信息，单击“提交”按钮即可，如下图所示。本书的作者和编辑会对您提交的勘误进行审核，确认并接受后，您将获赠异步社区的 100 积分。积分可用于在异步社区兑换优惠券、样书或奖品。

The screenshot shows a web form titled "提交勘误" (Submit Error Report) with three tabs: "详细信息" (Detailed Information), "写书评" (Write a Review), and "提交勘误" (Submit Error Report). The form contains three input fields: "页码:" (Page Number), "页内位置 (行数):" (Page Position (Line Number)), and "勘误印次:" (Error Report Count). Below these fields is a rich text editor with a toolbar containing icons for bold (B), italic (I), underline (U), strikethrough (ABC), bulleted list (•), numbered list (1-2-3), link (S), and unlink (S). In the bottom right corner of the form, there is a "字数统计" (Character Count) label and a "提交" (Submit) button.

扫码关注本书

扫描下方二维码，您将会在异步社区微信服务号中看到本书信息及相关的服务提示。



与我们联系

我们的联系邮箱是 contact@epubit.com.cn。

如果您对本书有任何疑问或建议，请您发邮件给我们，请在邮件标题中注明本书书名，以便我们更高效地做出反馈。

如果您有兴趣出版图书、录制教学视频，或者参与图书翻译、技术审校等工作，可以发邮件给我们；有意出版图书的作者也可以到异步社区在线提交投稿（直接访问 www.epubit.com/selfpublish/submission 即可）。

如果您是学校、培训机构或企业用户，想批量购买本书或异步社区出版的其他图书，也可以发邮件给我们。

如果您在网上发现有针对异步社区出品图书的各种形式的盗版行为，包括对图书全部或部分内容的非授权传播，请您将怀疑有侵权行为的链接发邮件给我们。您的这一举动是对作者权益的保护，也是我们持续为您提供有价值的内容的动力之源。

关于异步社区和异步图书

“异步社区”是人民邮电出版社旗下 IT 专业图书社区，致力于出版精品 IT 技术图书和相关学习产品，为译者提供优质出版服务。异步社区创办于 2015 年 8 月，提供大量精品 IT 技术图书和电子书，以及高品质技术文章和视频课程。更多详情请访问异步社区官网 <https://www.epubit.com>。

“异步图书”是由异步社区编辑团队策划出版的精品 IT 专业图书的品牌，依托于人民邮电出版社近 30 年的计算机图书出版积累和专业编辑团队，相关图书在封面上印有异步图书的 LOGO。异步图书的出版领域包括软件开发、大数据、人工智能、软件测试、前端、网络技术等。



异步社区



微信服务号

目 录

第 1 章 图像的获取和表示	1	2.2.3 图像增强	31
1.1 图像传感器技术	1	2.2.4 为特征提取准备图像	32
1.1.1 传感器材料	2	2.3 图像处理方法分类	36
1.1.2 传感器光电二极管元件	3	2.3.1 点运算	36
1.1.3 传感器配置: 马赛克、 Foveon 和 BSI	3	2.3.2 直线运算	36
1.1.4 动态范围、噪声和超分辨率	4	2.3.3 区域运算	37
1.1.5 传感器处理	5	2.3.4 算法	37
1.1.6 去马赛克	5	2.3.5 数据转换	37
1.1.7 坏像素校正	5	2.4 色彩学	37
1.1.8 色彩和光照校正	6	2.4.1 色彩管理系统概述	38
1.1.9 几何校正	6	2.4.2 光源、白点、黑点和中性轴	38
1.2 照相机和计算成像	6	2.4.3 设备颜色模型	39
1.2.1 计算成像概述	7	2.4.4 色彩空间与色彩感知	39
1.2.2 单像素可计算相机	7	2.4.5 色域映射与渲染的目标	40
1.2.3 二维可计算照相机	8	2.4.6 色彩增强的实际考虑	41
1.2.4 三维深度的照相机系统	9	2.4.7 色彩的准确度与精度	41
1.3 三维深度处理	18	2.5 空间滤波	41
1.3.1 方法概述	18	2.5.1 卷积滤波与检测	41
1.3.2 深度感知和处理中存在的 问题	18	2.5.2 核滤波与形状选择	43
1.3.3 单目深度处理	23	2.5.3 点滤波	44
1.4 三维表示: 体元、深度图、 网格和点云	26	2.5.4 噪声与伪像滤波	45
1.5 总结	27	2.5.5 积分图与方框滤波器	46
1.6 习题	27	2.6 边缘检测器	46
第 2 章 图像预处理	29	2.6.1 核集合	47
2.1 图像处理概述	29	2.6.2 Canny 检测器	48
2.2 图像预处理要解决的问题	29	2.7 变换滤波、Fourier 变换及其他	48
2.2.1 计算机视觉的流程和 图像预处理	30	2.7.1 Fourier 变换	48
2.2.2 图像校正	31	2.7.2 其他变换	50
		2.8 形态学与分割	51
		2.8.1 二值形态学	51
		2.8.2 灰度和彩色形态学	52

2.8.3	形态学优化和改进	53	3.4.7	Karhunen-Loeve 变换与 Hotelling 变换	93
2.8.4	欧氏距离映射	53	3.4.8	小波变换和 Gabor 滤波器	93
2.8.5	超像素分割	53	3.4.9	Hough 变换与 Radon 变换	95
2.8.6	深度图分割	54	3.5	总结	96
2.8.7	色彩分割	55	3.6	习题	96
2.9	阈值化	55	第 4 章	局部特征设计	97
2.9.1	全局阈值化	56	4.1	局部特征	97
2.9.2	局部阈值化	59	4.1.1	检测器、兴趣点、关键点、 锚点和特征点	98
2.10	总结	60	4.1.2	描述子、特征描述和 特征提取	98
2.11	习题	60	4.1.3	稀疏局部模式方法	98
第 3 章	全局特征和区域特征	63	4.2	局部特征属性	99
3.1	视觉特征的历史概述	63	4.2.1	选择特征描述子和兴趣点	99
3.1.1	全局度量、区域度量和局部 度量的核心思想	64	4.2.2	特征描述子和特征匹配	99
3.1.2	纹理分析	65	4.2.3	好特征的标准	99
3.1.3	统计方法	68	4.2.4	可重复性, 困难和 容易的查找	101
3.2	纹理区域度量	68	4.2.5	判别性与非判别性	101
3.2.1	边缘度量	69	4.2.6	相对位置和绝对位置	101
3.2.2	互相关性和自相关性	70	4.2.7	匹配代价和一致性	101
3.2.3	Fourier 谱、小波和基签名	71	4.3	距离函数	102
3.2.4	共生矩阵、Haralick 特征	71	4.3.1	距离函数的早期工作	102
3.2.5	Laws 纹理度量	78	4.3.2	欧氏或笛卡儿距离度量	103
3.2.6	LBP 局部二值模式	79	4.3.3	网格距离度量	104
3.2.7	动态纹理	80	4.3.4	基于统计学的差异性度量	105
3.3	统计区域度量	81	4.3.5	二值或布尔距离度量	106
3.3.1	图像矩特征	81	4.4	描述子的表示	107
3.3.2	点度量特征	81	4.4.1	坐标空间和复合空间	107
3.3.3	全局直方图	83	4.4.2	笛卡儿坐标	107
3.3.4	局部区域直方图	83	4.4.3	极坐标和对数极坐标	107
3.3.5	散点图、3D 直方图	84	4.4.4	径向坐标	107
3.3.6	多分辨率、多尺度直方图	85	4.4.5	球面坐标	108
3.3.7	径向直方图	87	4.4.6	Gauge 坐标	108
3.3.8	轮廓或边缘直方图	87	4.4.7	多元空间和多模数据	108
3.4	基空间度量	88	4.4.8	特征金字塔	109
3.4.1	Fourier 描述	90	4.5	描述子的密度	109
3.4.2	Walsh-Hadamard 变换	90	4.5.1	丢弃兴趣点和描述子	109
3.4.3	HAAR 变换	91	4.5.2	稠密与稀疏特征描述	110
3.4.4	斜变换	91			
3.4.5	Zernike 多项式	91			
3.4.6	导向滤波器	92			

4.6 描述子形状	110	第 5 章 特征描述属性的分类	141
4.6.1 关联性模板	111	5.1 一般的鲁棒性分类	143
4.6.2 块和形状	111	5.2 一般的视觉度量分类	146
4.6.3 对象多边形	113	5.3 特征度量评估	155
4.7 局部二值描述子与点对 模式	113	5.3.1 SIFT 的示例	156
4.7.1 FREAK 视网膜模式	114	5.3.2 LBP 的示例	156
4.7.2 BRISK 模式	115	5.3.3 形状因子的示例	157
4.7.3 ORB 和 BRIEF 模式	116	5.4 总结	158
4.8 描述子的判别性	116	5.5 习题	158
4.8.1 谱的判别性	117	第 6 章 兴趣点检测与特征描述子	159
4.8.2 区域、形状和模式的判别性	118	6.1 兴趣点调整	159
4.8.3 几何判别因素	118	6.2 兴趣点的概念	160
4.8.4 通过特征可视化来 评价判别性	119	6.3 兴趣点方法概述	162
4.8.5 精度与可跟踪性	121	6.3.1 Laplacian 和 LoG	163
4.8.6 精度优化、子区域重叠、 Gaussian 加权和池化	122	6.3.2 Moravac 角点检测器	163
4.8.7 亚像素精度	123	6.3.3 Harris 方法、Harris-Stephens、 Shi-Tomasi 和 Hessian 类型的 检测器	163
4.9 搜索策略与优化	123	6.3.4 Hessian 矩阵检测器和 Hessian-Laplace	164
4.9.1 密集搜索	124	6.3.5 Gaussian 差	164
4.9.2 网格搜索	124	6.3.6 显著性区域	164
4.9.3 多尺度金字塔搜索	124	6.3.7 SUSAN、Trajkovic-Hedly	165
4.9.4 尺度空间和图像金字塔	125	6.3.8 FAST	165
4.9.5 特征金字塔	126	6.3.9 局部曲率方法	166
4.9.6 稀疏预测搜索与跟踪	127	6.3.10 形态兴趣区域	167
4.9.7 跟踪区域限制搜寻	127	6.4 特征描述简介	167
4.9.8 分割限制搜索	127	6.4.1 局部二值描述子	168
4.9.9 深度或 Z 限制搜索	127	6.4.2 Census	173
4.10 计算机视觉、模型和结构	128	6.4.3 改进的 Census 变换	174
4.10.1 特征空间	128	6.4.4 BRIEF	174
4.10.2 对象模型	129	6.4.5 ORB	175
4.10.3 约束	130	6.4.6 BRISK	176
4.10.4 选择检测器和特征	131	6.4.7 FREAK	176
4.10.5 训练概述	131	6.5 谱描述子	177
4.10.6 特征和对象的分类	132	6.5.1 SIFT	177
4.10.7 特征学习、稀疏编码和 卷积网络	136	6.5.2 SIFT-PCA	181
4.11 总结	139	6.5.3 SIFT-GLOH	181
4.12 习题	139	6.5.4 SIFT-SIFER	182
		6.5.5 SIFT CS-LBP	182

6.5.6	ROOTSIFT	183	7.2	先前关于基准数据方面的工作： 艺术与科学	212
6.5.7	CenSurE 和 STAR	183	7.2.1	质量的一般度量	212
6.5.8	相关模板	185	7.2.2	算法性能的度量	212
6.5.9	HAAR 特征	186	7.2.3	Rosin 关于角点方面的工作	213
6.5.10	使用类 HAAR 特征的 Viola 和 Jones 算法	187	7.3	构造基准数据的关键问题	214
6.5.11	SURF	187	7.3.1	内容：采用、修改或创建	214
6.5.12	改进的 SURF 算法	189	7.3.2	可用的基准数据集	215
6.5.13	梯度直方图 (HOG) 及 改进方法	189	7.3.3	拟合基准数据的算法	215
6.5.14	PHOG 和相关方法	190	7.3.4	场景构成和标注	216
6.5.15	Daisy 和 O-Daisy	191	7.4	定义目标和预期	218
6.5.16	CARD	193	7.4.1	Mikolajczyk 和 Schmid 的 方法	218
6.5.17	具有鲁棒性的快速 特征匹配	194	7.4.2	开放式评价系统	219
6.5.18	RIFF 和 CHOG	195	7.4.3	极端情况和限制	219
6.5.19	链码直方图	196	7.4.4	兴趣点和特征	219
6.5.20	D-NETS	196	7.5	基准数据的鲁棒性准则	220
6.5.21	局部梯度模式	197	7.5.1	举例说明鲁棒性标准	220
6.5.22	局部相位量化	198	7.5.2	将鲁棒性标准用于 实际应用	221
6.6	基空间描述子	198	7.6	度量与基准数据配对	222
6.6.1	Fourier 描述子	199	7.6.1	兴趣点、特征和基准数据的 配对和优化	222
6.6.2	用其他基函数来构建 描述子	200	7.6.2	一般的视觉分类例子	223
6.6.3	稀疏编码方法	200	7.7	合成的特征字母表	224
6.7	多边形形状描述	200	7.7.1	合成数据集的目标	224
6.7.1	MSER 方法	201	7.7.2	合成兴趣点字母表	226
6.7.2	针对斑点和多边形的目标 形状度量	202	7.7.3	将合成字母表叠加到 真实图像上	228
6.7.3	形状上下文	204	7.8	总结	229
6.8	3D 和 4D 描述子	205	7.9	习题	230
6.8.1	3D HOG	206	第 8 章	可视流程及优化	231
6.8.2	HON 4D	206	8.1	阶段、操作和资源	231
6.8.3	3D SIFT	207	8.2	计算资源预算	233
6.9	总结	208	8.2.1	计算单元、ALU 和加速器	234
6.10	习题	208	8.2.2	能耗的使用	235
第 7 章	基准数据、内容、 度量和分析	210	8.2.3	内存的利用	235
7.1	基准数据	210	8.2.4	I/O 性能	238
			8.3	计算机视觉流程的实例	238
			8.3.1	汽车识别	239

8.3.2	人脸检测、情感识别和 年龄识别	244	21 世纪初	280	
8.3.3	图像分类	250	9.6.2	人工神经网络 (ANN) 分类	284
8.3.4	增强现实	254	9.7	特征学习概述	285
8.4	可选的加速方案	258	9.7.1	通过学习得到的 各类描述子	285
8.4.1	内存优化	258	9.7.2	层次特征学习	285
8.4.2	粗粒度并行	260	9.7.3	要学习多少特征	286
8.4.3	细粒度数据并行	261	9.7.4	深度神经网络的优势	286
8.4.4	高级指令集和加速器	263	9.7.5	特征编码的有效性	286
8.5	视觉算法的优化与调整	263	9.7.6	手工设计的特征与 深度学习	287
8.5.1	编译器优化与手工优化	264	9.7.7	特征学习的不变性和 鲁棒性	288
8.5.2	特征描述子改进、检测器和 距离函数	265	9.7.8	最好的特征和学习架构	288
8.5.3	Boxlets 与卷积加速	265	9.7.9	大数据、分析和计算机 视觉的统一	289
8.5.4	数据类型优化 (整数与 浮点)	265	9.7.10	关键技术的推动因素	291
8.6	优化资源	266	9.8	神经科学的概念	292
8.7	总结	266	9.8.1	生物学及其整体结构	293
第 9 章	特征学习的架构分类和 神经科学背景	267	9.8.2	难以找到统一的学习理论	294
9.1	计算机视觉中的神经 科学思想	268	9.8.3	人类视觉系统的架构	295
9.2	特征生成与特征学习	269	9.9	特征学习的结构分类	299
9.3	计算机视觉中所使用的 神经科学术语	269	9.9.1	架构拓扑	301
9.4	特征学习的分类	274	9.9.2	架构组件和层	302
9.4.1	卷积特征权重学习	275	9.10	总结	313
9.4.2	局部特征描述子学习	275	9.11	习题	313
9.4.3	基本特征的组合和 字典学习	275	第 10 章	特征学习和深度学习 架构概述	315
9.4.4	特征学习方法总结	276	10.1	架构概述	315
9.5	计算机视觉中的机器 学习模型	276	10.1.1	FNN 架构简介	316
9.5.1	专家系统	277	10.1.2	RNN 的结构简介	372
9.5.2	统计和数学分析方法	278	10.1.3	BFN 的结构简介	395
9.5.3	受神经科学启发的方法	278	10.2	集成方法	427
9.5.4	深度学习	278	10.3	深度神经网络的未来	429
9.6	机器学习和特征学习的 历史	280	10.3.1	增加最大深度——深度残差 学习	429
9.6.1	历史回顾: 20 世纪 40 年代至 21 世纪初	280	10.3.2	使用更简单的 MLP 来近似复杂 模型 (模型压缩)	430
			10.3.3	分类器的分解与重组	431

10.4 总结	432	附录 D 扩展 SDM 准则	474
10.5 习题	432	附录 E 视觉基因组模型 (VGM)	487
附录 A 合成特征分析	435	参考文献	508
附录 B 基准数据集概述	464	译后记	541
附录 C 成像和计算机视觉资源	470		

第 1 章 图像的获取和表示

计算机视觉的基础是图像。本章介绍的内容包括图像获取，图像处理和表示图像（计算成像的方法、二维成像方法和三维深度成像方法），传感器处理，对立体图像和单目多视图立体图像的景深（depth-field）处理，表面重构。本章会概括性地介绍一些主题，并给出参考文献以供感兴趣的读者进一步研究。对二维和三维成像有较好基础的读者也可从本章受益。

1.1 图像传感器技术

本节简要介绍图像传感器技术，它是理解图像如何构成，如何优化计算机视觉的图像质量而有效进行图像处理的基础。

典型的图像传感器有 CCD 单元（charge-coupled device，电荷耦合器件）或者标准 CMOS 单元（complementary meta-oxide semiconductor，互补金属氧化物半导体）。CCD 和 CMOS 传感器具有相似的特性，广泛应用于商业相机上。不过，现在多数传感器均使用 CMOS 单元，这主要出于制造方面的考虑。传感器和光学器件常常整合在一起用于制造晶片级照相机，这种照相机用在类似于生物学或显微镜学等领域，如图 1-1 所示。

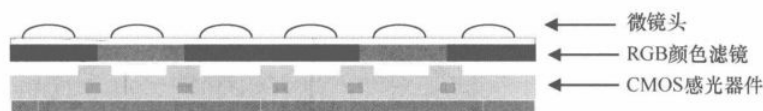


图 1-1 整合了光学器件和颜色过滤器的图像传感器的常用排列

图像传感器是为满足不同应用的特殊目标而设计的，它提供了不同级别的灵敏度和质量。可咨询制造商了解各种传感器的详细信息。比如，为了在硅基模和动态响应（用于实现发光强度和颜色检测）之间有一个最好的折中，对一个特定的半导体制造过程，需要优化每个光电二极管传感器单元的大小和组成成分。

对计算机视觉而言，采样理论的效果具有重要意义——比如，目标场景的像素范围就会用到 Nyquist 频率。传感器分辨率和光学器件可为每个像素提供足够的分辨率，从而使其对感兴趣的特征进行成像，因此有这样的结论：对感兴趣特征的采样（或成像）频率至少应该是重要像素（对感兴趣的特征而言）中最小像素的两倍。当然，对于成像精度而言，两倍的过采样（oversampling）仅仅是一个最低目标；在实际应用中，并不容易决定单像素宽度特征。

对于给定的应用，要取得最好的结果，需标定照相机，以便能在不同光照（lighting）、不同距离的情况下针对像素位深度（bit depth）确定传感器噪声和动态范围。为了能处理传感器对任何颜色通道所产生的噪声、非线性响应、检测和校正像素坏点、处理几何失真的建模，需要研究合适的传感器处理方法。如果采用的是一种简单标定方法，即在灰度、颜色、特征像素尺度等方面使用由细到粗的渐进测试方法，那么就需要设计恰当的传感器处理方法。第2章会讨论一系列用于传感器处理的图像处理方法，这里还是先介绍传感器材料。

1.1.1 传感器材料

硅制图像传感器应用最广，当然也会使用其他材料，比如在工业和军事应用中会用镓（Ga）来覆盖比硅更长的红外波长。不同照相机的图像传感器的分辨率会有所不同，比如采用单像素的光电晶体管照相机；在工业应用上使用的一维线扫描相机；普通照相机上采用的二维矩形阵列；高分辨率成像上采用的球形阵列。（本章最后会介绍传感器配置和照相机配置。）

通常采用基于硅材料¹的 CCD、CMOS、BSI 和 Foveon 方式来制造普通成像传感器（本章稍后会讨论）。尽管基于硅材料的图像传感器具有非线性的光谱响应曲线，并且这会很好地感知光谱的近红外部分，但它对蓝色、紫色和近紫外部分的感知会欠佳（如图 1-2 所示）。如果直接读取原始数据，并离散化为数字像素时，那么要特别注意硅的光谱响应曲线。传感器制造商在设计上做了补偿，但根据应用标定照相机并设计传感器处理方法时，应该考虑传感器的颜色响应问题。

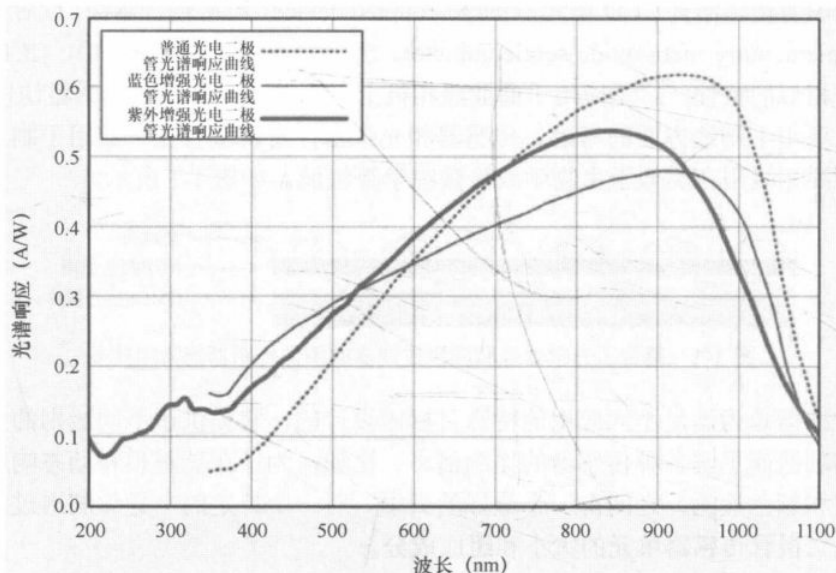


图 1-2 几种硅光电二极管的典型光谱响应。注意，在 900nm 附近的近红外范围内具有高的灵敏度，而在 400~700nm 的可见光范围内具有非线性灵敏度。由于标准的硅响应的缘故，从照相机中去掉 IR 滤波器会增加近红外的灵敏度（光谱数据图像的使用已获得 OSI 光电股份有限公司的许可）

¹ 也简称为“硅基”。